

Families of 2-weights of some particular graphs

Agnese Baldisserri, Elena Rubei

Abstract

Let $\mathcal{G} = (G, w)$ be a positive-weighted graph, that is a graph G endowed with a function w from the edge set of G to the set of positive real numbers; for any distinct vertices i, j , we define $D_{i,j}(\mathcal{G})$ to be the weight of the path in G joining i and j with minimum weight. In this paper we fix a particular class of graphs and we give a criterion to establish whether, given a family of positive real numbers $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$, there exists a positive-weighted graph $\mathcal{G} = (G, w)$ in the class we have fixed, with vertex set equal to $\{1, \dots, n\}$ and such that $D_I(\mathcal{G}) = D_I$ for any $I \in \binom{\{1, \dots, n\}}{2}$. In particular, the classes of graphs we consider are the following: snakes, caterpillars, polygons, bipartite graphs, complete graphs, planar graphs.

1 Introduction

For any graph G , let $E(G)$, $V(G)$ and $L(G)$ be respectively the set of the edges, the set of the vertices and the set of the leaves of G . A **weighted graph** $\mathcal{G} = (G, w)$ is a graph G endowed with a function $w : E(G) \rightarrow \mathbb{R}$. For any edge e , the real number $w(e)$ is called the weight of the edge and, for any subgraph H of G , we define $w(H)$ to be the sum of the weights of the edges of H . We say that \mathcal{G} is **positive-weighted** if all the weights of the edges are positive.

Definition 1. Let $\mathcal{G} = (G, w)$ be a positive-weighted graph; for any distinct $i, j \in V(G)$ we define

$$D_{\{i,j\}}(\mathcal{G}) = \min\{w(H) \mid H \text{ a connected subgraph of } G \text{ such that } V(H) \ni i, j\}.$$

More simply, we denote $D_{\{i,j\}}(\mathcal{G})$ by $D_{i,j}(\mathcal{G})$. We say that a connected subgraph H **realizes** $D_{i,j}(\mathcal{G})$ if $V(H) \ni i, j$ and $w(H) = D_{i,j}(\mathcal{G})$. Obviously such a subgraph must be a path with endpoints i and j . We call the $D_{i,j}(\mathcal{G})$ the **2-weights** of \mathcal{G} .

Throughout the paper we will consider only simple finite connected graphs. Observe that in the case G is a tree, $D_{i,j}(\mathcal{G})$ is the weight of the unique path joining i and j .

If S is a subset of $V(G)$, the 2-weights $D_{i,j}(\mathcal{G})$ with $i, j \in S$ give a vector in $\mathbb{R}^{\binom{S}{2}}$. This vector is called **2-dissimilarity vector** of (\mathcal{G}, S) . Equivalently, we can speak of the **family of the 2-weights** of (\mathcal{G}, S) .

We can wonder when a family of real numbers is the family of the 2-weights of some positive-weighted graph and of some subset of the set of its vertices. If S is a finite set of cardinality greater than 2, we say that a family of real numbers $\{D_I\}_{I \in \binom{S}{2}}$ is **p-graphlike** if there exist a positive-weighted

2010 Mathematical Subject Classification: 05C05, 05C12, 05C22

Key words: weighted graphs, 2-weights

graph $\mathcal{G} = (G, w)$ and a subset S of the set of its vertices such that $D_I(\mathcal{G}) = D_I$ for any 2-subset I of S . If the graph is a positive-weighted tree $\mathcal{T} = (T, w)$ we say that the family is p-treelike.

Weighted graphs have applications in several disciplines, such as biology, psychology, archeology, engineering. Phylogenetic trees are positive-weighted trees whose vertices represent species and the weight of an edge is given by how much the DNA sequences of the species represented by the vertices of the edge differ; also archeologists represent evolutions of manuscripts by positive-weighted trees. There is a wide literature concerning graphlike dissimilarity families and treelike dissimilarity families, in particular concerning methods to reconstruct weighted trees from their dissimilarity families; these methods are used by biologists to reconstruct phylogenetic trees. See for example [8], [16] for overviews on phylogenetic trees. Weighted graphs can represent hydraulic webs or railway webs where the weight of an edge is given by the length or the cost (or the difference between the earnings and the cost) of the line represented by that edge.

The first contribution to the characterization of graphlike families of numbers dates back to 1965 and it is due to Hakimi and Yau, see [11]:

Theorem 2. (Hakimi-Yau) *A family of positive real numbers $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ is p-graphlike if and only if the D_I satisfy the triangle inequalities, i.e. if and only if $D_{i,j} \leq D_{i,k} + D_{k,j}$ for any distinct $i, j, k \in [n]$.*

In the same years, also a criterion for a metric on a finite set to be p-treelike was established, see [7], [17], [18]:

Theorem 3. (Buneman-SimoesPereira-Zaretskii) *Let $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ be a family of positive real numbers satisfying the triangle inequalities. It is p-treelike if and only if the D_I satisfy the so-called four-point condition, i.e., for all distinct $i, j, k, h \in \{1, \dots, n\}$, the maximum of*

$$\{D_{i,j} + D_{k,h}, D_{i,k} + D_{j,h}, D_{i,h} + D_{k,j}\}$$

is attained at least twice.

Also the case of not necessarily nonnegative weights has been studied, see [10] for graphs and [6] for trees.

Finally we want to mention that recently k -weights of weighted graphs for $k \geq 3$ have been introduced and studied; in particular there are some results concerning the characterization of families of k -weights, see for instance [1], [2], [12], [13], [14], [15]. The study of k -weights for $k \geq 3$ is motivated by the fact that they are more reliable statistically than 2-weights and so the reconstruction of weighted trees from them can be more accurate than the reconstruction from 2-weights.

In this paper, we fix a particular class of graphs, we consider a family of positive real numbers $\{D_I\}_{I \in \binom{\{1, \dots, n\}}{2}}$ and we give a criterion to establish whether there exists a positive-weighted graph $\mathcal{G} = (G, w)$ in the class we have fixed, with $V(G) = \{1, \dots, n\}$ and such that $D_I(\mathcal{G}) = D_I$ for any $I \in \binom{\{1, \dots, n\}}{2}$. In particular the classes we consider are the following: snakes, caterpillars, polygons, bipartite graphs, complete graphs, planar graphs.

2 Some definitions and some remarks

Notation 4. • Let $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x > 0\}$.

- Throughout the paper let $n \in \mathbb{N} - \{0, 1\}$ and let $[n] = \{1, \dots, n\}$.

- If G is a graph, for any $v, v' \in V(G)$, let $e(v, v')$ denote the edge joining v and v' .
- For any graph G , let $V^i(G)$ be the set of the vertices of G of degree i and let $V^{\geq i}(G) = \cup_{j \geq i} V^j(G)$.
- For any graph G , we say that an edge is **pendant** if it is incident to a leaf.
- If T is a tree, for any $v, v' \in V(T)$, let $p(v, v')$ denote the unique path joining v and v' .
- Throughout the paper, the word “tree” will denote a finite tree and the word “graph” will denote a finite connected graph.
- Let $n \in \mathbb{N} - \{0, 1\}$. For any family of real numbers $\{D_I\}_{I \in \binom{[n]}{2}}$, we denote $D_{\{i,j\}}$ by $D_{i,j}$.

Definition 5. • A **snake** is a tree with only 2 leaves.

- We say that a tree C is a **caterpillar** if there is a path S such that $V(S) = V^{\geq 2}(C)$. We call S the **spine** of the caterpillar.
- A graph P with n vertices is a **polygon** if we can rename the vertices by i_1, \dots, i_n in such way that $E(P) = \{e(i_1, i_2), \dots, e(i_{n-1}, i_n), e(i_n, i_1)\}$.
- A graph G is **complete** if $E(G)$ contains the edge $e(i, j)$ for any $i, j \in V(G)$. The complete graph with n vertices is usually denoted by K_n .
- a graph B is a **bipartite graph** on two subsets X and Y of $V(B)$ if:
 - $X \cap Y = \emptyset$;
 - $X \cup Y = V(B)$;
 - $E(B) \subset \{e(x, y) \mid x \in X, y \in Y\}$.
- A **bipartite graph** B on two subsets X and Y of $V(B)$ is **complete** if $E(B)$ contains the edge $e(i, j)$ for any $i \in X, j \in Y$. The complete bipartite graph on two sets one of cardinality m and one of cardinality n is usually denoted by $K_{m,n}$.

Remark 6. Let B be a graph and let $X, Y, P, Q \subset V(B)$ such that B is a bipartite graph on X and Y and is a bipartite graph on P and Q ; then we can easily show that $X = P$ and $Y = Q$ or vice versa.

Definition 7. Let $\mathcal{G} = (G, w)$ a positive-weighted graph, we say that an edge e of G is **useful** if there exist $i, j \in V(G)$ such that all the paths realizing the 2-weight $D_{i,j}(\mathcal{G})$ contain the edge e . We say that an edge e is **useless** if it is not useful, that is, if all the 2-weights of the graph are realized by at least a path which do not contain e . Finally, we say that a graph \mathcal{G} is **pruned** if all its edges are useful.

Definition 8. Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ . We say that the family is

- **snakelike** if there exists a positive-weighted snake $\mathcal{S} = (S, w)$ with $V(S) = [n]$ such that $D_{i,j}(\mathcal{S}) = D_{i,j}$ for any $i, j \in [n]$,
- **caterpillarlike** if there exists a positive-weighted caterpillar $\mathcal{C} = (C, w)$ with $V(C) = [n]$ such that $D_{i,j}(\mathcal{C}) = D_{i,j}$ for any $i, j \in [n]$,
- **polygonlike** if there exists a positive-weighted polygon $\mathcal{P} = (P, w)$ with $V(P) = [n]$ such that $D_{i,j}(\mathcal{P}) = D_{i,j}$ for any $i, j \in [n]$,

- **co-graphlike** if there exists a positive-weighted complete graph $\mathcal{G} = (G, w)$ with $V(G) = [n]$ such that $D_{i,j}(\mathcal{G}) = D_{i,j}$ for any $i, j \in [n]$ and G is pruned,
- **bigraphlike** on two subsets X and Y of $[n]$ if there exists a positive-weighted bipartite graph $\mathcal{B} = (B, w)$ with $V(B) = [n]$ on X, Y such that $D_{i,j}(\mathcal{B}) = D_{i,j}$ for any $i, j \in [n]$,
- **co-bigraphlike** if there exists a pruned positive-weighted complete bipartite graph $\mathcal{B} = (B, w)$ with $V(B) = [n]$ such that $D_{i,j}(\mathcal{B}) = D_{i,j}$ for any $i, j \in [n]$,
- **planargraphlike** if there exists a positive-weighted planar graph $\mathcal{G} = (G, w)$ with $V(G) = [n]$ such that $D_{i,j}(\mathcal{G}) = D_{i,j}$ for any $i, j \in [n]$.

To be precise we should say “*p*-snakelike, *p*-caterpillarlike...” to point out that we are considering positive-weighted graphs, but, since we will consider only positive-weighted graphs and so no confusion can arise, for simplicity we will omit the letter “*p*”.

Remark 9. Let $\mathcal{G} = (G, w)$ be a positive-weighted graph with $V(G) = [n]$ and let $i, j \in [n]$; if $D_{i,j}(\mathcal{G})$ is realized by a path H in G , then, for any $k, t \in V(H)$, the 2-weight $D_{k,t}(\mathcal{G})$ is realized by the path in H with endpoints k and t .

Proof. Suppose, contrary to our claim, that there exist $k, t \in V(H)$ such that any path realizing $D_{k,t}(\mathcal{G})$ is not contained in H . Call J one of the paths realizing $D_{k,t}(\mathcal{G})$ and call H' a path joining k with t contained in H . Then we would have:

$$w(H') > D_{k,t}(\mathcal{G}) = w(J); \quad (1)$$

moreover,

$$w(H) = D_{i,j}(\mathcal{G}) \leq w((H \setminus H') \cup J) \leq w(H \setminus H') + w(J);$$

thus

$$w(H') \leq w(J),$$

which is absurd because it contradicts (1). \square

Definition 10. Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ satisfying the triangle inequalities; we say that an element of the family $D_{i,j}$ is **indecomposable** if $D_{i,j} < D_{i,z} + D_{z,j}$ for any $z \in [n] \setminus \{i, j\}$.

Remark 11. Let $\mathcal{G} = (G, w)$ be a positive-weighted graph such that $V(G) = [n]$. For any $i, j \in [n]$, the 2-weight $D_{i,j}(\mathcal{G})$ is indecomposable if and only if $E(G)$ contains the edge $e(i, j)$ and $e(i, j)$ is useful. In this case we have that the 2-weight $D_{i,j}(\mathcal{G})$ is realized only by the edge $e(i, j)$ and, in particular, $D_{i,j}(\mathcal{G}) = w(e(i, j))$.

Proof. Suppose that $D_{i,j}(\mathcal{G})$ is indecomposable; if it were realized by a path joining i with j different from $e(i, j)$, it would contain another vertex $z \in [n] \setminus \{i, j\}$, then, by Remark 9, we would have that $D_{i,j}(\mathcal{G}) = D_{i,z}(\mathcal{G}) + D_{z,j}(\mathcal{G})$, which is absurd. So $D_{i,j}(\mathcal{G})$ can be realized only by $e(i, j)$, thus $e(i, j) \in E(G)$ and $e(i, j)$ useful. Conversely, suppose to have a useful edge $e(i, j) \in E(G)$: by definition, there exist two vertices $a, b \in V(G)$ such that all the paths realizing $D_{a,b}(\mathcal{G})$ contain $e(i, j)$. Then the 2-weight $D_{i,j}(\mathcal{G})$ is realized by the edge $e(i, j)$ (by Remark 9) and it can be realized only by the edge $e(i, j)$, so it is indecomposable. \square

3 Snakes and caterpillars

In this section we give a characterization of snakelike families in \mathbb{R}_+ and a characterization of caterpillarlike ones.

Theorem 12. *Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ and let $x, y \in [n]$ such that $D_{x,y} = \max_{i,j \in [n]} \{D_{i,j}\}$; the family is snakelike if and only if $D_{i,j} = |D_{i,x} - D_{j,x}|$ for any distinct $i, j \in [n] \setminus \{x\}$.*

Proof. \implies Very easy to prove.

\impliedby First note that $D_{a,x} \neq D_{b,x}$ for any distinct $a, b \in [n] \setminus \{x\}$; otherwise we would have:

$$D_{a,b} = |D_{a,x} - D_{b,x}| = 0,$$

which is absurd because, by assumption, the elements of the family are positive. Let us denote the elements of $[n] \setminus \{x, y\}$ by i_1, i_2, \dots, i_{n-2} in such a way that

$$D_{i_j,x} < D_{i_{j+1},x}$$

for any $j = 1, \dots, n-3$ and let $\mathcal{S} = (S, w)$ be the positive-weighted snake defined as follows (see Figure 1): let S be the snake with $V(S) = [n]$ and $E(S) = \{e(x, i_1), e(i_1, i_2), \dots, e(i_{n-2}, y)\}$ and define the weights of $e(x, i_1), e(i_1, i_2), \dots, e(i_{n-2}, y)$ to be, respectively, $D_{i_1,x}, D_{i_2,x} - D_{i_1,x}, D_{i_3,x} - D_{i_2,x}, \dots, D_{y,x} - D_{i_{n-2},x}$.

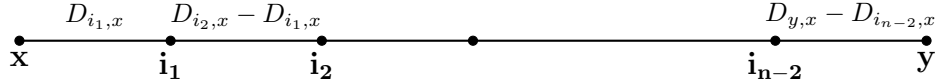


Figure 1: a positive-weighted snake realizing the family $\{D_{i,j}\}$

We have to check that $D_{i,j}(\mathcal{S}) = D_{i,j}$ for any $i, j \in [n]$:

- $D_{i,x}(\mathcal{S}) = D_{i,x}$ for any $i \in [n] \setminus \{x\}$ by construction;
- $D_{i,j}(\mathcal{S}) = |D_{i,x}(\mathcal{S}) - D_{j,x}(\mathcal{S})| = |D_{i,x} - D_{j,x}| = D_{i,j}$ for any $i, j \in [n] \setminus \{x\}$, where the last equality holds by assumption.

□

Before studying caterpillarlike families, we introduce a definition and we state a theorem that will be useful later:

Definition 13. *Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ . We say that the family $\{D_I\}$ is a **median family** if, for any $a, b, c \in [n]$, there exists a unique element $m \in [n]$ such that*

$$D_{i,j} = D_{i,m} + D_{j,m}$$

for any distinct $i, j \in \{a, b, c\}$.

Observe that a median family satisfies the triangle inequalities. The theorem below, probably well-known to experts, was suggested to us by an anonymous referee in October 2014; later we have found it also in [9]; we defer to [5] for a shorter proof.

Theorem 14. Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ . There exists a positive-weighted tree $\mathcal{T} = (T, w)$, with $V(T) = [n]$, such that $D_I(\mathcal{T}) = D_I$ for all $I \in \binom{[n]}{2}$ if and only if the four-point condition holds and the family $\{D_I\}_I$ is median.

Now, consider a positive-weighted caterpillar $\mathcal{C} = (C, w)$ with $V(C) = [n]$.

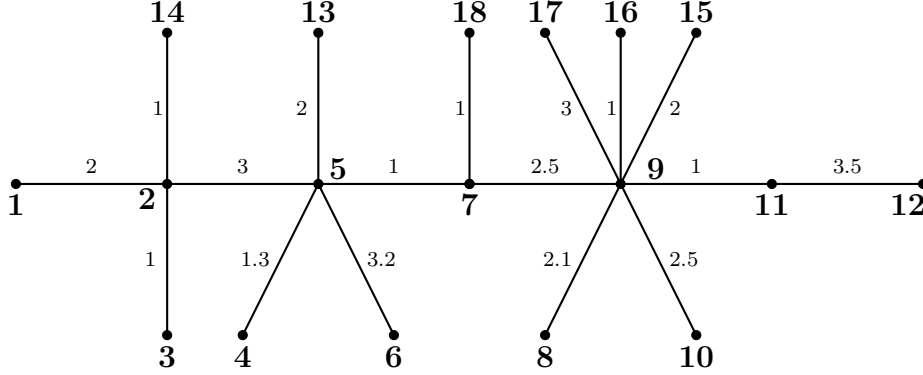


Figure 2: a positive-weighted caterpillar $\mathcal{C} = (C, w)$ with $V(C) = [18]$

Given a vertex $x \in V(C)$, we can define

$$t_x = \frac{1}{2} \min_{z, y \in V(C) \setminus \{x\}} \{D_{x,y}(C) + D_{x,z}(C) - D_{y,z}(C)\};$$

it is easy to show that, if $x \in L(C)$, then t_x is the weight of the pendant edge associated to x and that $t_x = 0$ if and only if $x \notin L(C)$, that is, x belongs to the spine of C .

Remark 15. Let $\mathcal{C} = (C, w)$ be a positive-weighted caterpillar with $V(C) = [n]$ and let $x_1, x_2 \in V(C)$ be such that $p(x_1, x_2)$ is the spine of C . Let X_1 (respectively X_2) be the set of the leaves of C adjacent to x_1 (respectively x_2) (for example, in Figure 2, we have that $\{x_1, x_2\} = \{2, 11\}$ and, if we take for instance $x_1 = 2$ and $x_2 = 11$, we have that $X_1 = \{1, 3, 14\}$ and $X_2 = \{12\}$). If we consider two vertices $a, b \in V(C)$ such that

$$D_{a,b}(C) - t_a - t_b = \max_{i,j \in [n]} \{D_{i,j}(C) - t_i - t_j\},$$

we have that $a \in X_1 \cup \{x_1\}$ and $b \in X_2 \cup \{x_2\}$ or vice versa.

Proof. It is sufficient to note that, for any $i, j \in V(C)$,

$$D_{i,j}(C) - t_i - t_j = w(p(\bar{i}, \bar{j})),$$

where \bar{i} is defined as follows: it is equal to i if $i \notin L(C)$, while it is the vertex adjacent to i if $i \in L(C)$; analogously for \bar{j} . \square

Remark 16. Let $\mathcal{C} = (C, w)$ a positive-weighted tree with $V(C) = [n]$; call a, b two vertices of C such that

$$D_{a,b}(C) - t_a - t_b = \max_{i,j \in [n]} \{D_{i,j}(C) - t_i - t_j\}.$$

The tree C is a caterpillar if and only if for any $i, j \in [n] \setminus \{a, b\}$ we have that

$$D_{a,b}(C) + D_{i,j}(C) \geq \max\{D_{a,i}(C) + D_{b,j}(C), D_{a,j}(C) + D_{b,i}(C)\}. \quad (2)$$

Proof. If C is a caterpillar, then, using Remark 15 it is easy to check that (2) holds for any $i, j \in [n] \setminus \{a, b\}$. Now, suppose that C is not a caterpillar, then there must be a vertex c with degree grater than 1 which is not in $p(a, b)$.

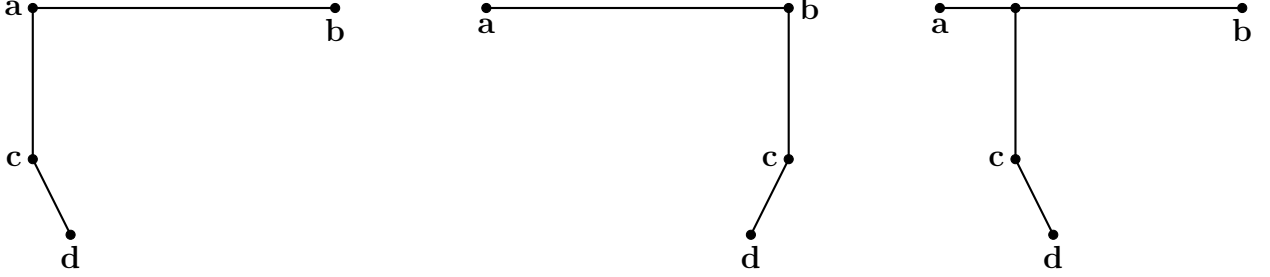


Figure 3: cases (1), (2) and (3)

We have three cases:

1. $p(a, c) \cap p(a, b) = \{a\}$;
2. $p(a, c) \cap p(a, b) = p(a, b)$;
3. $\{a\} \subsetneq p(a, c) \cap p(a, b) \subsetneq p(a, b)$.

Here we study the third case, the other ones are analogous. Call d a vertex such that $p(c, d) \cap p(c, a) = \{c\}$ (see Figure 3), we have that

$$D_{a,b}(\mathcal{C}) + D_{c,d}(\mathcal{C}) < D_{a,c}(\mathcal{C}) + D_{b,d}(\mathcal{C})$$

and

$$D_{a,b}(\mathcal{C}) + D_{c,d}(\mathcal{C}) < D_{a,d}(\mathcal{C}) + D_{b,c}(\mathcal{C}),$$

which is absurd. □

Now we are ready to give a characterization of caterpillarlike families of positive real numbers:

Theorem 17. *Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ . Call a, b two elements of $[n]$ such that*

$$D_{a,b} - t_a - t_b = \max_{i,j \in [n]} \{D_{i,j} - t_i - t_j\};$$

the family is caterpillarlike if and only if the following conditions hold:

- (i) *the family satisfies the four-point condition,*
- (ii) *the family is median,*
- (iii) $D_{a,b} + D_{i,j} \geq \max\{D_{a,i} + D_{b,j}, D_{a,j} + D_{b,i}\}$ *for any $i, j \in [n] \setminus \{a, b\}$.*

Proof. Suppose that the family $\{D_I\}_I$ is caterpillarlike, then there exists a positive-weighted caterpillar $\mathcal{C} = (C, w)$ with $V(C) = [n]$ realizing the family. Obviously the family must satisfy conditions (i) and (ii) and, by Remark 16, also condition (iii) holds.

On the other hand, suppose to have a family of positive real numbers satisfying conditions (i), (ii) and (iii). Since the family satisfies the four-point condition and is median, by Theorem 14 there exists a positive-weighted tree $\mathcal{C} = (C, w)$ with $V(C) = [n]$ realizing the family. Moreover, by condition (iii) and Remark 16, C is a caterpillar, as we wanted to prove. \square

Finally, we want to observe that, if we have a family of numbers in \mathbb{R}_+ which is caterpillarlike and we call a, b two elements of $[n]$ such that $D_{a,b} - t_a - t_b = \max_{i,j \in [n]} \{D_{i,j} - t_i - t_j\}$, then it is easy to construct a positive-weighted caterpillar realizing the family: it is sufficient to draw a path of length $D_{a,b}$ with endpoints a and b ; for any $i \in [n] \setminus \{a, b\}$, if $t_i > 0$ attach a pendant edge with weight t_i and leaf i to a point of the path which has distance from a equal to $D_{a,i} - t_i$; if $t_i = 0$, call i a vertex on the path which has distance from a equal to $D_{a,i}$.

4 Polygons

Let $\mathcal{P} = (P, w)$ be a positive-weighted polygon such that $V(P) = [n]$.

Observe that in case \mathcal{P} is not pruned, there is at most one useless edge e . So, if we delete e , we obtain a positive-weighted snake $\tilde{\mathcal{P}} = (\tilde{P}, \tilde{w})$ with $V(\tilde{P}) = [n]$ and with the same family of 2-weights.

Suppose now that \mathcal{P} is pruned: by Remark 11, for any $i \in V(P)$, the vertices x and y adjacent to i are exactly the ones such that $D_{i,x}(\mathcal{P})$ and $D_{i,y}(\mathcal{P})$ are indecomposable so it is possible to recover the order of the vertices of the polygon starting from the 2-weights.

Definition 18. Let P be a polygon with $[n]$ as vertex set. We say that the vertex set is sequentially ordered if i and $i + 1$ are adjacent for any $i \in [n - 1]$ and n and 1 are adjacent.

Definition 19. Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ satisfying the triangle inequalities and such that, for any $i \in [n]$, there exist exactly two elements $x, y \in [n] \setminus \{i\}$ for which $D_{i,x}$ and $D_{i,y}$ are indecomposable. We can rename the elements of $[n]$ according the following algorithm:

rename 1 and 2 two elements of $[n]$ such that $D_{1,2} = \min\{D_{i,j}\}$ and define $H = \{1, 2\}$. Observe that $D_{1,2}$ must be indecomposable. Rename 3 the unique element in $[n] \setminus \{1, 2\}$ such that $D_{2,3}$ is indecomposable and put 3 in H . Recursively, call $i + 1$ the unique element in $[n] \setminus \{i - 1, i\}$ such that $D_{i,i+1}$ is indecomposable. If $i + 1 \in H$ stop the algorithm, otherwise put $i + 1$ in H .

Theorem 20. Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ satisfying the triangle inequalities; there exists a pruned positive-weighted polygon $\mathcal{P} = (P, w)$ with $V(P) = [n]$ realizing the family if and only if the following conditions hold:

- (i) for any $i \in [n]$ there are exactly two elements $x, y \in [n]$ such that $D_{i,x}$ and $D_{i,y}$ are indecomposable;
- (ii) if H is the set described in Definition 19, then the cardinality of H is n ;
- (iii) if the elements of $[n]$ are renamed as in Definition 19, then, for any $a < b \in [n]$, we have that

$$D_{a,b} = \min \left\{ \sum_{i=a}^{b-1} D_{i,i+1}, \sum_{i=b}^{n-1} D_{i,i+1} + D_{1,n} + \sum_{i=1}^{a-1} D_{i,i+1} \right\}. \quad (3)$$

Proof. Suppose that there exists a pruned positive-weighted polygon $\mathcal{P} = (P, w)$ with $V(P) = [n]$ such that $D_{i,j}(\mathcal{P}) = D_{i,j}$ for any $i, j \in [n]$. It is easy to check that conditions (i) and (ii) hold. Moreover, if we rename the vertices as in Definition 19, the vertex set is sequentially ordered. Since \mathcal{P} is pruned, for any $i \in [n-1]$ the 2-weight $D_{i,i+1}(\mathcal{P})$ is realized by $e(i, i+1)$ and the 2-weight $D_{1,n}(\mathcal{P})$ is realized by $e(1, n)$ (see Remark 11). Obviously, for any two vertices $a, b \in [n]$, with $a < b$, a subgraph realizing the 2-weight $D_{a,b}(\mathcal{P})$ is a path with endpoints a and b and in the polygon there are exactly two different paths with endpoints a and b . Their weights are the numbers at the second member of (3), so we get condition (iii).

On the other hand, let $\{D_I\}_I$ be a family of positive real numbers satisfying conditions (i), (ii) and (iii). By condition (i) and (ii) we can rename all the elements of $[n]$ as in Definition 19. Let $\mathcal{P} = (P, w)$ be the positive-weighted polygon with $V(P) = [n]$, with the vertex set sequentially ordered, and such that $w(e(i, i+1)) = D_{i,i+1}$ for any $i \in [n-1]$ and $w(e(1, n)) = D_{1,n}$ (see Figure 4).

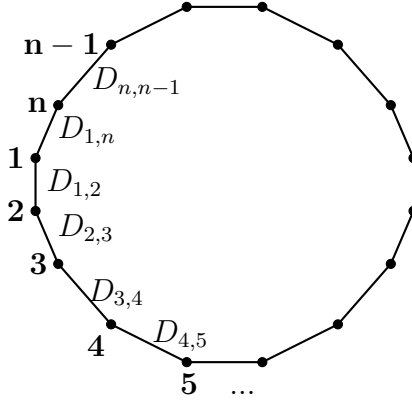


Figure 4:

We have to prove that $D_{a,b}(\mathcal{P}) = D_{a,b}$ for any $a, b \in [n]$; obviously a subgraph realizing $D_{a,b}(\mathcal{P})$ is a path with endpoints a and b and in the polygon there are exactly two different paths with endpoints a and b . By the definition of \mathcal{P} , their weights are the two numbers at the second member of (3), so we have that

$$D_{a,b}(\mathcal{P}) = \min \left\{ \sum_{i=a}^{b-1} D_{i,i+1}, \sum_{i=b}^{n-1} D_{i,i+1} + D_{1,n} + \sum_{i=1}^{a-1} D_{i,i+1} \right\} = D_{a,b},$$

where the last equality holds by (3). Observe that \mathcal{P} is pruned, in fact, if an edge $e(a, b)$ (with a and b adjacent vertices) were useless, then $D_{a,b}(\mathcal{P})$ would not be indecomposable, which is absurd because we have constructed \mathcal{P} in such a way that two vertices are adjacent if and only if $D_{a,b}$ is indecomposable. \square

Now we can give a characterization of the families of positive real numbers that are polygonlike:

Theorem 21. *A family of positive real numbers $\{D_I\}_{I \in \binom{[n]}{2}}$ satisfying the triangle inequalities is polygonlike if and only if either it is snakelike or it satisfies conditions (i), (ii) and (iii) of Theorem 20.*

Proof. Suppose there exists a positive-weighted polygon $\mathcal{P} = (P, w)$ with $V(P) = [n]$ realizing the family; if \mathcal{P} is pruned, then by Theorem 20 the family must satisfy conditions (i), (ii) and (iii). If \mathcal{P} is not pruned, then we can delete the unique useless edge and we obtain a positive-weighted snake realizing the family, so the family is snakelike.

Conversely, suppose there exists a positive-weighted snake $\mathcal{S} = (S, w)$ with $V(S) = [n]$ realizing the family. If i, j are the endpoints of the snake, we can add to the snake an edge $e(i, j)$ with weight any real number greater than or equal to $D_{i,j}$: it is easy to check that the positive-weighted polygon with n vertices we have obtained realizes the family $\{D_I\}_I$, so the family is also polygonlike. Finally, if the family satisfies conditions (i), (ii) and (iii) of Theorem 20, it is polygonlike by Theorem 20. \square

5 Complete graphs and bipartite graphs

An immediate consequence of Remark 11 is the following characterization of the co-graphlike families of 2-weights:

Remark 22. *Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ satisfying the triangle inequalities; the family is co-graphlike if and only if $D_{i,j}$ is indecomposable for any $i, j \in [n]$.*

Proof. Suppose there exists a pruned positive-weighted complete graph $\mathcal{G} = (G, w)$ with $V(G) = [n]$ realizing the family; thus $e(i, j)$ is useful for any $i, j \in [n]$; so, by Remark 11, $D_{i,j}(\mathcal{G})$ is indecomposable for any $i, j \in [n]$. On the other hand, suppose $D_{i,j}$ is indecomposable for any $i, j \in [n]$; by Theorem 2, there exists a positive-weighted graph $\mathcal{G} = (G, w)$ with $V(G) = [n]$ realizing the family; moreover, since $D_{i,j}$ is indecomposable for any $i, j \in [n]$, then, by Remark 11, we have that $e(i, j) \in E(G)$ and $e(i, j)$ is useful for any $i, j \in [n]$. \square

Now we want to characterize bigraphlike families of positive real numbers; first of all, given a positive-weighted bipartite graph $\mathcal{G} = (G, w)$ on $X, Y \subset V(G)$, we show that it is possible to recover X and Y from the family of 2-weights of \mathcal{G} .

Proposition 23. *Let $\mathcal{B} = (B, w)$ be a positive-weighted bipartite graph on X and Y with $V(B) = [n]$; let $x \in X$ and $y \in Y$; then:*

$$X = \{x\} \cup \left\{ i \in [n] \setminus \{x, y\} \mid \exists j_1, \dots, j_t \in [n] \text{ with } t \text{ odd such that} \right.$$

$$\left. D_{x,i}(\mathcal{B}) = D_{x,j_1}(\mathcal{B}) + D_{j_1,j_2}(\mathcal{B}) + \dots + D_{j_t,i}(\mathcal{B}) \text{ and the elements of the sum are indecomposable} \right\} \quad (4)$$

and

$$Y = \left\{ i \in [n] \setminus \{x\} \mid \text{either } D_{x,i} \text{ is indecomposable or } \exists j_1, \dots, j_t \in [n] \text{ with } t \text{ even such that} \right.$$

$$\left. D_{x,i}(\mathcal{B}) = D_{x,j_1}(\mathcal{B}) + D_{j_1,j_2}(\mathcal{B}) + \dots + D_{j_t,i}(\mathcal{B}) \text{ and the elements of the sum are indecomposable} \right\}. \quad (5)$$

Proof. Let us prove (4); the other equality can be proved analogously. Call R the second member of (4); we want to prove that $X = R$.

- $X \subset R$: let $i \in X \setminus \{x\}$; observe that $D_{x,i}(\mathcal{B})$ is not indecomposable: otherwise by Remark 11, we would have $e(x, i) \in E(B)$, which is absurd; so we can write $D_{x,i}(\mathcal{B})$ as

$$D_{x,j_1}(\mathcal{B}) + D_{j_1,j_2}(\mathcal{B}) + \dots + D_{j_t,i}(\mathcal{B})$$

for some j_1, \dots, j_t with $D_{x,j_1}(\mathcal{B}), D_{j_1,j_2}(\mathcal{B}), \dots, D_{j_t,i}(\mathcal{B})$ indecomposable. By Remark 11, the 2-weights $D_{x,j_1}(\mathcal{B}), D_{j_1,j_2}(\mathcal{B}), \dots, D_{j_t,i}(\mathcal{B})$ are realized respectively by $e(x, j_1), e(j_1, j_2), \dots, e(j_t, i)$; thus the path given by the union of these edges realizes $D_{x,i}(\mathcal{B})$ and, since $x, i \in X$, we have that t is necessarily odd.

- $R \subset X$: if $i \in R$ then there exist $j_1, \dots, j_t \in [n]$ with t odd such that $D_{x,i}(\mathcal{B}) = D_{x,j_1}(\mathcal{B}) + D_{j_1,j_2}(\mathcal{B}) + \dots + D_{j_t,i}(\mathcal{B})$ and the elements of the sum are indecomposable. By Remark 11, the 2-weights $D_{x,j_1}(\mathcal{B}), D_{j_1,j_2}(\mathcal{B}), \dots, D_{j_t,i}(\mathcal{B})$ are realized respectively only by the edges $e(x, j_1), e(j_1, j_2), \dots, e(j_t, i)$, which implies that $D_{x,i}(\mathcal{B})$ is realized by the path given by these edges; so, since t is odd, $i \in X$.

□

Remark 24. Let $\mathcal{B} = (B, w)$ be a positive-weighted bipartite graph on X and Y with $V(B) = [n]$. Let $x, y \in [n]$ be such that

$$D_{x,y}(\mathcal{B}) = \min_{1 \leq i < j \leq n} \{D_{i,j}(\mathcal{B})\};$$

hence, obviously, $D_{x,y}(\mathcal{B})$ is indecomposable, and then, by Remark 9, $e(x, y) \in E(B)$ and $D_{x,y}(\mathcal{B})$ is realized only by the path with unique edge $e(x, y)$.

Now we are ready to give a characterization of the families of positive real numbers that are bigraphlike:

Theorem 25. Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ satisfying the triangle inequalities and let $x, y \in [n]$ such that $D_{x,y} = \min_{1 \leq i < j \leq n} \{D_{i,j}\}$; define

$$X = \{x\} \cup \left\{ i \in [n] \setminus \{x, y\} \mid \exists j_1, \dots, j_t \in [n] \text{ with } t \text{ odd such that} \right.$$

$$\left. D_{x,i} = D_{x,j_1} + D_{j_1,j_2} + \dots + D_{j_t,i} \text{ and the elements of the sum are indecomposable} \right\}$$

and

$$Y = \left\{ i \in [n] \setminus \{x\} \mid \text{either } D_{x,i} \text{ is indecomposable or } \exists j_1, \dots, j_t \in [n] \text{ with } t \text{ even such that} \right.$$

$$\left. D_{x,i} = D_{x,j_1} + D_{j_1,j_2} + \dots + D_{j_t,i} \text{ and the elements of the sum are indecomposable} \right\}.$$

The family $\{D_I\}_I$ is bigraphlike if and only if the following conditions hold:

- (1) $X \cap Y = \emptyset$
- (2) for any $a, b \in X$ (respectively Y), there exists $z \in Y$ (respectively X) such that:

$$D_{a,b} = D_{a,z} + D_{z,b}.$$

Proof. Suppose there exist two subset X and Y of $[n]$ and a positive-weighted bipartite graph $\mathcal{B} = (B, w)$ on X' and Y' with $V(B) = [n]$ realizing the family. By Proposition 23 we have that $X = X'$ and $Y = Y'$ (or vice versa), so $X \cap Y = \emptyset$. Let $a, b \in X$; a path realizing $D_{a,b}(\mathcal{B})$ must contain a vertex $z \in Y$, so, by Remark 9, we have that $D_{a,b}(\mathcal{B}) = D_{a,z}(\mathcal{B}) + D_{z,b}(\mathcal{B})$. If both a and b are elements of Y , the proof is analogous.

Now, suppose that $\{D_I\}_I$ satisfies (1) and (2). Let $\mathcal{B} = (B, w)$ be the positive-weighted bipartite graph on X and Y such that:

- $V(G) = [n]$;
- $E(G) = \{(a, b) \mid a \in X, b \in Y\}$;
- $w(e(a, b)) = D_{a,b}$ for any $a \in X$ and $b \in Y$.

We want to prove that $D_{a,b}(\mathcal{B}) = D_{a,b}$ for any $a, b \in [n]$. Let p be a path realizing $D_{a,b}(\mathcal{B})$ and let $j_1, \dots, j_t \in [n]$ be such that p is given by $e(a, j_1), e(j_1, j_2), \dots, e(j_t, b)$; then we have that

$$D_{a,b}(\mathcal{B}) = w(e(a, j_1)) + w(e(j_1, j_2)) + \dots + w(e(j_t, b)) = D_{a,j_1} + D_{j_1,j_2} + \dots + D_{j_t,b} \geq D_{a,b}, \quad (6)$$

where the last inequality follows from the triangle inequalities.

If $a \in X$ and $b \in Y$ (or vice versa), then

$$D_{a,b}(\mathcal{B}) \leq w(e(a, b)) = D_{a,b}, \quad (7)$$

so, from (6) and (7), we get $D_{a,b}(\mathcal{B}) = D_{a,b}$.

If both a and b are in X (if they are in Y , we can argue analogously), by assumption, there exists $z \in Y$ such that $D_{a,z} + D_{b,z} = D_{a,b}$; we have that

$$D_{a,b}(\mathcal{B}) \leq D_{a,z} + D_{b,z} = D_{a,b}, \quad (8)$$

where the inequality holds because the path given by $e(a, z)$ and $e(z, b)$ contains a and b as vertices and its weight is equal to $D_{a,z} + D_{b,z}$; so, from (6) and (8), we get also in this case $D_{a,b}(\mathcal{B}) = D_{a,b}$. \square

Finally we give also a characterization of co-bigraphlike families of positive real numbers:

Remark 26. Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ which is bigraphlike on $X, Y \subset [n]$. The family is co-bigraphlike on X and Y if and only if $D_{i,j}$ is indecomposable for any $i \in X, j \in Y$.

Proof. Let $\mathcal{B} = (B, w)$ be a positive-weighted complete bipartite graph on X and Y , with $V(B) = [n]$, realizing the family. If it is pruned, then $e(i, j)$ is useful for any $i \in X, j \in Y$; so, by Remark 11, $D_{i,j}(\mathcal{B})$ is indecomposable for any $i \in X, j \in Y$. On the other hand, if $D_{i,j}$ is indecomposable for any $i \in X$ and $j \in Y$, then, by Remark 11, $e(i, j) \in E(B)$ and $e(i, j)$ is useful for any $i \in X, j \in Y$. \square

6 Planar graphs

Definition 27. Let G be a graph and let $e(u, v)$ be an edge of G . We say that a graph G' is obtained from G by a subdivision of the edge $e(u, v)$ if $V(G')$ is the union of $V(G)$ and a new vertex z and $E(G')$ is $E(G) - \{e(u, v)\} \cup \{e(u, z), e(z, v)\}$. We say that a graph G' is a subdivision of a graph G if it is the graph resulting from the subdivision of some edges in G .

Theorem 28. (Kuratowski) *A finite graph is planar if and only if it does not contain a subgraph that is a subdivision of K_5 or of $K_{3,3}$.*

Definition 29. *Let G be a subdivision of K_5 . We say that a vertex of G is a **true vertex** if it is a vertex of K_5 . We call **verges** of G the paths that are subdivisions of the edges of K_5 .*

Proposition 30. *Let $\mathcal{G} = (G, w)$ be a positive-weighted graph with $V(G) = [n]$. Suppose it is pruned. Let us denote $D_{i,j}(\mathcal{G})$ by $D_{i,j}$ for any $i, j \in [n]$.*

(i) *G contains a subdivision of $K_5 \iff$ there exists $Q \in \binom{[n]}{5}$ such that for any distinct $a, b \in Q$, either $D_{a,b}$ is indecomposable or there exists a sequence (x_1, \dots, x_r) in $[n] - Q$ (depending on $\{a, b\}$) such that $D_{a,x_1}, \dots, D_{x_r,b}$ are indecomposable and, if $\{a, b\} \neq \{a', b'\}$, the sequence of $\{a, b\}$ does not intersect the sequence of $\{a', b'\}$.*

(ii) *G contains a subdivision of $K_{3,3} \iff$ there exist disjoint $A, B \in \binom{[n]}{3}$ such that for any $a \in A$ and $b \in B$, either $D_{a,b}$ is indecomposable or there exists a sequence (x_1, \dots, x_r) in $[n] - A - B$ (depending on $\{a, b\}$) such that $D_{a,x_1}, \dots, D_{x_r,b}$ are indecomposable and, if $\{a, b\} \neq \{a', b'\}$, the sequence of $\{a, b\}$ does not intersect the sequence of $\{a', b'\}$.*

Proof. Let us prove (i) (the proof of (ii) is analogous).

\Leftarrow By Remark 11, if $D_{i,j}$ is indecomposable, then $e(i, j) \in E(G)$. For any $a, b \in Q$, let $c_{a,b}$ be the following path: the path given only by the edge $e(a, b)$ if $D_{a,b}$ is indecomposable, the path given by the edges $e(a, x_1), e(x_1, x_2), \dots, e(x_r, b)$ if (x_1, \dots, x_r) is a sequence as in the statement of the proposition.

The union of the paths $c_{a,b}$ for $a, b \in Q$ gives a subgraph that is a subdivision of K_5 .

\Rightarrow Let G' be a subdivision of K_5 in G . Let Q be the set of the true vertices of G' . Since \mathcal{G} is pruned, every edge is useful, in particular, for any $x, y \in [n]$ such that $e(x, y)$ is in a verge of G' , we have that $e(x, y)$ is useful, so, by Remark 11, the 2-weight $D_{x,y}$ is indecomposable and then we get our statement. \square

Theorem 31. *Let $\{D_I\}_{I \in \binom{[n]}{2}}$ be a family in \mathbb{R}_+ . It is planargraphlike if and only if the following conditions hold:*

(a) *the family satisfies the triangle inequalities;*

(b) *there does not exist $Q \in \binom{[n]}{5}$ such that, for any distinct $a, b \in Q$, either $D_{a,b}$ is indecomposable or there exists a sequence (x_1, \dots, x_r) in $[n] - Q$ (depending on $\{a, b\}$) such that $D_{a,x_1}, \dots, D_{x_r,b}$ are indecomposable and, if $\{a, b\} \neq \{a', b'\}$, the sequence of $\{a, b\}$ does not intersect the sequence of $\{a', b'\}$;*

(c) *there do not exist disjoint $A, B \in \binom{[n]}{3}$ such that, for any $a \in A$ and $b \in B$, either $D_{a,b}$ is indecomposable or there exists a sequence (x_1, \dots, x_r) in $[n] - A - B$ (depending on $\{a, b\}$) such that $D_{a,x_1}, \dots, D_{x_r,b}$ are indecomposable and, if $\{a, b\} \neq \{a', b'\}$, the sequence of $\{a, b\}$ does not intersect the sequence of $\{a', b'\}$.*

Proof. \Leftarrow Let \mathcal{G} be a positive-weighted planar graph realizing the family. By eliminating a useless edge, then another one and so on, we get a pruned positive-weighted planar graph realizing the family. So we can conclude by Proposition 30 and Theorem 28.

\Rightarrow By condition (a) and Theorem 2, there exists a positive-weighted graph \mathcal{G} realizing the family; by eliminating a useless edge, then another one and so on, we get a pruned positive-weighted planar graph realizing the family; by conditions (b) and (c) and using Proposition 30 and Theorem 28, we can conclude that it is planar. \square

Acknowledgements. This work was supported by the National Group for Algebraic and Geometric Structures, and their Applications (GNSAGA-INdAM). The first author was supported by Ente Cassa di Risparmio di Firenze.

References

- [1] A. Baldisserri, E. Rubei *On graphlike k -dissimilarity vectors*, Ann. Comb., 18 (3) 356-381 (2014)
- [2] A. Baldisserri, E. Rubei *Families of multiweights and pseudostars*, to appear on Advances in Applied Mathematics.
- [3] A. Baldisserri, E. Rubei *Treelike families of multiweights*, arXiv:1404.6799, v3
- [4] A. Baldisserri, E. Rubei *A characterization of dissimilarity families of trees* arXiv:1512.08496
- [5] A. Baldisserri, E. Rubei *A characterization of graphlike families of multiweights*
- [6] H-J Bandelt, M.A. Steel *Symmetric matrices representable by weighted trees over a cancellative abelian monoid*. SIAM J. Discrete Math. 8 (1995), no. 4, 517–525
- [7] P. Buneman *A note on the metric properties of trees*, Journal of Combinatorial Theory Ser. B 17 (1974), 48-50
- [8] A. Dress, K. T. Huber, J. Koolen, V. Moulton, A. Spillner, Basic phylogenetic combinatorics. Cambridge University Press, Cambridge, 2012
- [9] M. Hayamizu, K. Fukumizu *On minimum spanning tree-like metric spaces* arXiv: 1505.06145
- [10] S.L. Hakimi, A.N. Patrinos *The distance matrix of a graph and its tree realization*, Quart. Appl. Math. 30 (1972/73), 255-269
- [11] S.L. Hakimi, S.S. Yau, *Distance matrix of a graph and its realizability*, Quart. Appl. Math. 22 (1965), 305-317
- [12] S. Herrmann, K. Huber, V. Moulton, A. Spillner, *Recognizing treelike k -dissimilarities*, J. Classification 29 (2012), no. 3, 321-340
- [13] B. Iriarte Giraldo *Dissimilarity vectors of trees are contained in the tropical Grassmannian*, Electron. J. Combin. 17 (2010), no. 1
- [14] D. Levy, R. Yoshida, L. Pachter *Beyond pairwise distances: neighbor joining with phylogenetic diversity estimates*, Mol. Biol. Evol., 23 (2006), 491–498.
- [15] L. Pachter, D. Speyer *Reconstructing trees from subtree weights*, Appl. Math. Lett. 17 (2004), no. 6, 615–621
- [16] C. Semple, M. Steel, Phylogenetics. Oxford University Press, Oxford, 2003
- [17] J.M.S. Simoes Pereira *A Note on the Tree Realizability of a distance matrix*, J. Combinatorial Theory 6 (1969), 303-310
- [18] K.A. Zaretskii *Constructing trees from the set of distances between pendant vertices*, Uspehi Matematicheskikh Nauk. 20 (1965), 90-92